

An archaeobacterial promoter sequence assigned by RNA polymerase binding experiments

MICHAEL THOMM

Lehrstuhl für Mikrobiologie, Universität Regensburg, Universitätsstr. 31, 8400 Regensburg, Federal Republic of Germany

GÜNTER WICH

Lehrstuhl für Mikrobiologie, Universität München, Maria Ward Straße 1a, 8000 München 19, Federal Republic of Germany

JAMES W. BROWN

Department of Microbiology, Ohio State University, Columbus, OH 43210, U.S.A.

GERHARD FREY

Lehrstuhl für Mikrobiologie, Universität Regensburg, Universitätsstr. 31, 8400 Regensburg, Federal Republic of Germany

AND

BRUCE A. SHERF AND GREGORY S. BECKLER

Department of Microbiology, Ohio State University, Columbus, OH 43210, U.S.A.

THOMM, M., WICH, G., BROWN, J. W., FREY, G., SHERF, B. A., and BECKLER, G. S. 1989. An archaeobacterial promoter sequence assigned by RNA polymerase binding experiments. *Can. J. Microbiol.* **35**: 30–35.

To identify an archaeobacterial promoter sequence, nuclease protection studies with the purified RNA polymerase of *Methanococcus vannielii* were performed. The enzyme binds specifically both at protein-encoding (*hisA* and methyl CoM reductase, component C) and tRNA–rRNA genes. The binding region of the RNA polymerase extends from 30 base pairs (bp) upstream (–30) to 20 bp downstream (+20) from the *in vivo* transcription start site. This finding indicates that the archaeobacterial enzyme recognizes promoters without transacting transcription factors. The DNA segment protected from nuclease digestion by bound RNA polymerase contains an octanucleotide sequence centered at –25, which is conserved between the protein-encoding and the stable RNA genes. According to the specific binding of the enzyme to only DNA-fragments harbouring this motif, we propose the sequence TTTATATA as the major recognition signal of the *Methanococcus vannielii* RNA polymerase. Comparison of this motif with published archaeobacterial DNA sequences revealed the presence of homologous sequences at the same location upstream of 36 genes. We therefore consider the overall consensus TTTA_ATATA as a general element of promoters in archaeobacteria. In spite of the specific binding of the enzyme, most preparations of the *Methanococcus vannielii* RNA polymerase are unable to initiate transcription at the correct sites *in vitro*. Here we present first evidence for the possible existence of a transcription factor conferring the ability to the enzyme to initiate and terminate transcription specifically *in vitro*.

Key words: promoter, footprint, TATA box, RNA polymerase, transcription.

THOMM, M., WICH, G., BROWN, J. W., FREY, G., SHERF, B. A., et BECKLER, G. S. 1989. An archaeobacterial promoter sequence assigned by RNA polymerase binding experiments. *Can. J. Microbiol.* **35** : 30–35.

Dans le but d'identifier une séquence archaébactérienne de promoteur, des études sur la protection contre la nucléase ont été entreprises avec de l'ARN polymérase purifiée de *Methanococcus vannielii*. Cette enzyme se lie spécifiquement à la fois aux gènes qui encodent la protéine (*hisA* et méthyle CoM réductase, composant C) et aux gènes ARN_t–ARN_r. La région d'attachement de l'ARN polymérase s'étend depuis 30 bp en amont (–30) à 20 bp en aval (+20) du site de début de transcription *in vivo*. Cette découverte indique que l'enzyme archaébactérienne reconnaît les promoteurs sans l'intervention de facteurs de transcription. Le segment d'ADN protégé contre la digestion par la nucléase, grâce à l'ARN polymérase liée, contient une séquence de huit nucléotides qui est centrée à –25; cette séquence est conservée entre les gènes qui encodent la protéine et les gènes d'ARN stable. D'après l'attachement spécifique de l'enzyme aux seuls fragments d'ADN qui contiennent ce motif, nous proposons comme signal de reconnaissance principal de l'ARN polymérase de *Methanococcus vannielii* la séquence TTTATATA. Une comparaison de ce motif avec d'autres séquences d'ADN archaébactériens qui ont fait l'objet de publications révèle la présence de séquences homologues à la même localisation en amont de 36 gènes. Nous proposons donc l'organisation d'ensemble TTTA_ATATA comme un élément commun aux promoteurs chez les archaébactéries. Malgré l'attachement spécifique de cette enzyme, la plupart des préparations d'ARN polymérase de *Methanococcus vannielii* ne réussissent pas, *in vitro*, à amorcer la transcription dans les sites appropriés. Nous présentons ici la première évidence de l'existence possible d'un facteur de transcription qui confère à l'enzyme la capacité d'initier et de terminer la transcription, particulièrement *in vitro*.

Mots clés : promoteur, empreinte, boîte de TATA, transcription, ARN polymérase.

[Traduit par la revue]

Introduction

The first step of gene expression involves recognition of DNA sequences necessary for initiation of transcription. Within the classical prokaryotes, the eubacteria (Woese and Fox 1977), the recognition and binding site of RNA polymerase has been designated as promoter (Jacob et al. 1964). Typical promoters of *Escherichia coli* are characterized by two sequences that are conserved upstream of most protein-

encoding and all tRNA–rRNA genes. Both the consensus sequences, 5'-TATAATG-3' (the Pribnow box) located 10 bp upstream from the transcription start site (–10) and 5'-TTGACA-3' at –35, have been identified as promoter elements by nuclease protection experiments (Schaller et al. 1975; Pribnow 1975; Siebenlist et al. 1980). In eubacterial RNA polymerases (subunit composition $\beta'\beta\alpha_2\sigma$) one of the four subunits, the Sigma factor, is essential for promoter

recognition (Burgess et al. 1969). This is corroborated by the finding that promoters with recognition sequences different from the canonical consensus are expressed by a complex of $\beta'\beta\alpha_2$ with alternative Sigma factors (Hoopes and McClure 1987).

The purified RNA polymerases of eukaryotes are unable to bind specifically to a promoter sequence. In these cells, promoter recognition is mediated by transcription factors (Dyana and Tjian 1985). Different DNA sequences have been identified as binding sites for the various transcription factors involved in the expression of eukaryotic rRNA, protein, and tRNA genes that are expressed by the RNA polymerases I, II, and III, respectively. However, most promoters of protein-encoding genes are preceded by an AT-rich heptanucleotide sequence (the TATA box) located at -25 (Corden et al. 1980).

In archaebacteria, until very recently, both the transcription signals and the mechanism of promoter recognition were unknown. The major reason of this was the lack of specificity of most archaebacterial RNA polymerase preparations. Our approach to identify a promoter sequence in archaebacteria was to investigate binding of the purified RNA polymerase of *Methanococcus vannielii* to homologous genes. This enzyme offers some advantages compared with the RNA polymerases from extreme halophiles and sulphur-dependent archaebacteria that have been purified earlier (Zillig et al. 1982); the temperature and salt optima of its enzymatic activity are compatible with conditions used for enzymatic footprinting studies. Although from an extreme anaerobic organism, this enzyme is stable for the short incubation times necessary for nuclease protection and *in vitro* transcription assays even in the presence of oxygen. Here, we infer a general archaebacterial promoter sequence from previous footprinting studies (Thomm and Wich 1988; Brown et al. 1988; Thomm et al. 1988). Furthermore, we provide first evidence for the existence of an archaebacterial transcription factor.

Materials and methods

Preparation of $\gamma^{32}\text{P}$ -labelled ribonucleoside-5'-triphosphates

End labelling of ADP, GDP, CDP, and UDP with [^{32}P] orthophosphate was performed using the Gamma Prep-system (Promega Biotech) according to the recommendation of the manufacturer.

Synthesis and isolation of the *in vitro* RNA

The conditions for *in vitro* transcription were as described previously (Thomm and Stetter 1985). The RNA polymerase of *M. vannielii* was incubated at 30°C. For hybridization experiments with $\gamma^{32}\text{P}$ -labelled ribonucleoside-5'-triphosphates, 100 μCi (1 Ci = 37 GBq.) were used.

Southern hybridization experiments

The Southern hybridization of *in vitro* RNA was performed as described previously (Thomm and Stetter 1985).

Analysis of RNA products

Labelled RNA was isolated and separated by electrophoresis in 6% polyacrylamide urea gels as described by Kingston and Chamberlin (1981).

Results and discussion

A TATA box at -25 as a recognition signal of the *M. vannielii* RNA polymerase

To identify promoter signals in archaebacteria, binding of the RNA polymerase of *M. vannielii* to both protein-encoding (*hisA* and methyl CoM reductase (*mcr*) and tRNA-rRNA genes was studied. Specificity of binding was investigated by

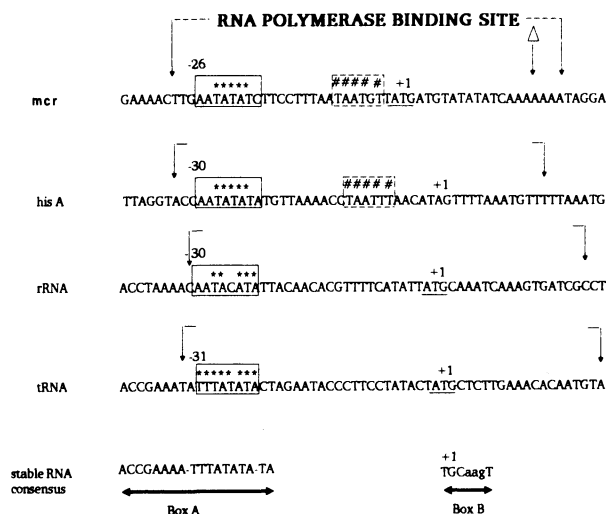


FIG. 1. DNA sequence of the RNA polymerase binding site at four genes of *Methanococcus vannielii*. The limits of the binding sites determined by exonuclease-III footprinting are labelled by vertical arrows. Nucleotides homologous to the octanucleotide sequence TTTATATA of the box-A motif (Wich et al. 1986) are labelled by asterisks. Homologous sequences at the RNA initiation sites are underlined. Δ Indicates the 3' boundary of a deletion clone. Sequences conserved between the protein-encoding genes (*hisA*, *mcr*; hatched box) are labelled by #.

various standard techniques, e.g., filter binding assays (Hinkle and Chamberlin 1972) and DNase-I (Galas and Schmitz 1978) and exonuclease-III footprinting (Shalloway et al. 1980). Specific binding of the highly purified enzyme to all four archaebacterial genes occurred (Thomm and Wich 1988; Brown et al. 1988; Thomm et al. 1988). On the contrary, no footprint was obtained with DNA fragments harbouring strong *E. coli* promoters (Thomm and Wich 1988) or intragenetic sequences of archaebacterial genes (M. Thomm and B. Sherf, unpublished). On all genes investigated, the enzyme binds exclusively to the *in vivo* transcription start site. The binding region extends from -30 to +20 (Fig. 1). Thus, apparently this archaebacterial enzyme is able to recognize promoters without additional factors like the *E. coli* RNA polymerase. This suggests that all sequences necessary for promoter recognition are contained within the binding region of the enzyme. Sequences within this region that are conserved between these genes are with high probability the recognition signals of the RNA polymerase. The protein-encoding genes (*mcr* and *hisA*) harbour at -25, an AT-rich region. This sequence shows homology to a part of the box-A motif (Wich et al. 1986), which has been suggested as a promoter element specific for stable RNA genes of *Methanococcus* (Fig. 1). Interestingly, only this part of the box-A sequence is contained within the binding region of the tRNA and rRNA genes (Fig. 1). Thus, 8 bp of the box-A consensus are conserved between protein-encoding and stable RNA genes. According to the specific binding of the enzyme, this sequence seems to represent the major recognition signal of the RNA polymerase. The box-A motif upstream of the binding site might be important for the regulation of expression of stable RNA genes. As it is absent at the protein-encoding genes (Fig. 1), where also specific binding of the enzyme occurs, this part apparently is not necessary for RNA polymerase binding.

A hexanucleotide sequence at -10 was found to be common to the *mcr* locus and the *hisA* gene and might represent a

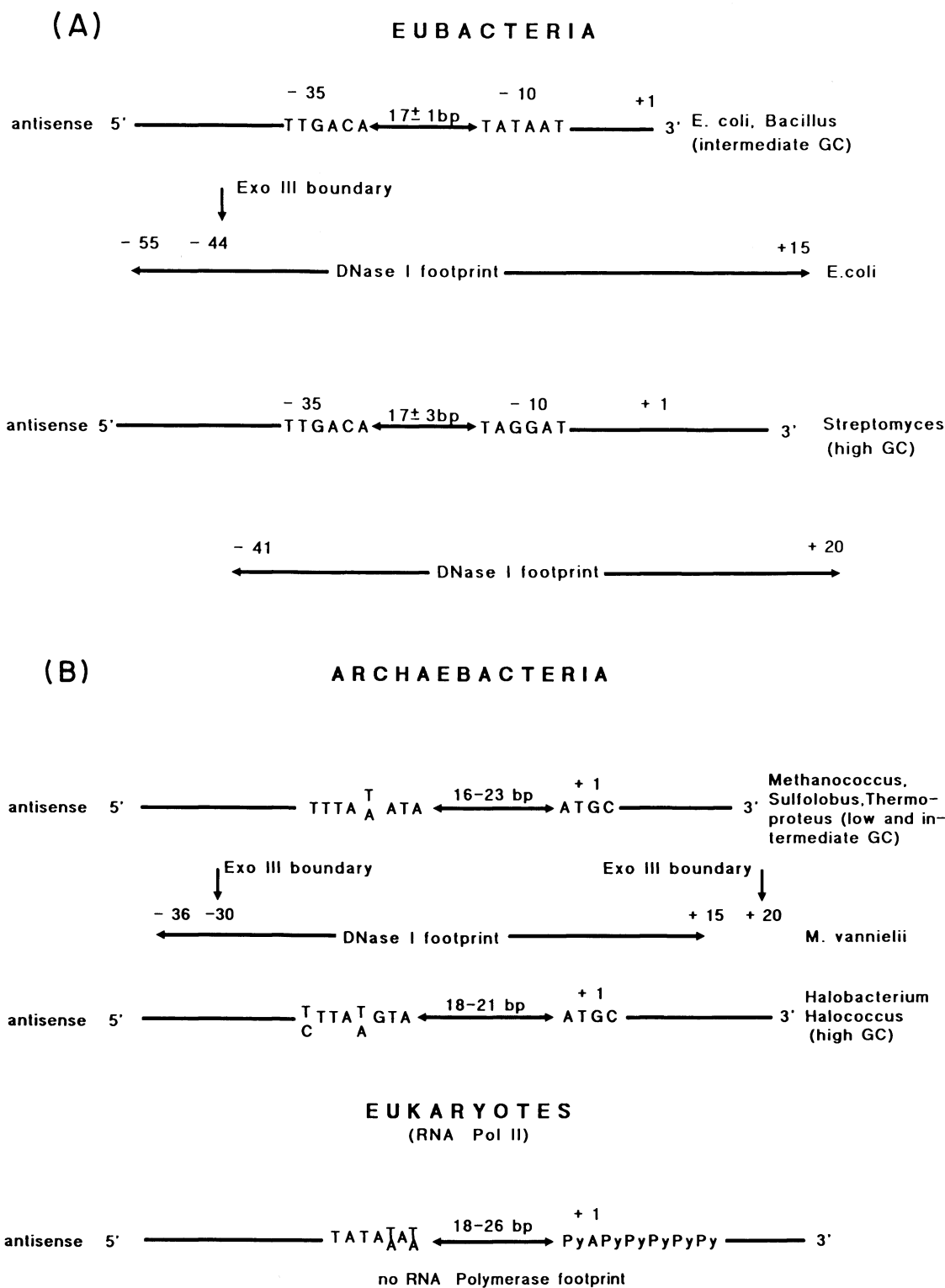


FIG. 2. Comparison of transcription signals and RNA polymerase binding sites in eubacteria, archaeobacteria, and eukaryotes. Exo, exonuclease; Pol, polymerase.

promoter sequence specific for protein-encoding genes of *Methanococcus* (Fig. 1, hatched box). Specific binding of the enzyme to the *mcr* gene was also observed when the region downstream from +16 was removed (Thomm et al. 1988; labelled with the symbol \triangle in Fig. 1). This indicates that

sequences downstream from this nucleotide are not essential for promoter recognition.

The sequence ATGC at the transcription start site is conserved between the *mcr* and the stable RNA genes (Fig. 1). This sequence might be a second archaeobacterial promoter ele-

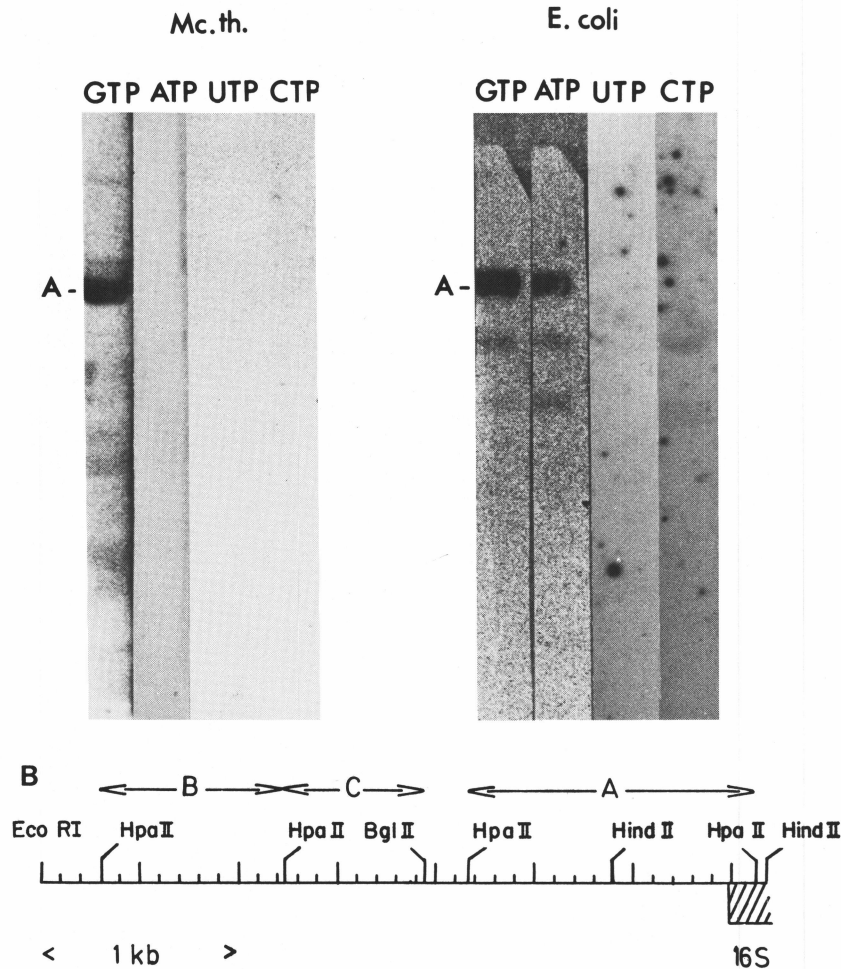


FIG. 3. Southern hybridization of *in vitro* RNA labelled with γ - ^{32}P -ribonucleoside-5'-triphosphates. The recombinant plasmid pMV15 Δ 3 (Thomm and Stetter 1985) harbouring the promoter of an rRNA operon of *M. vannielii* was used as a template for the *M. thermolithotrophicus* and *E. coli* RNA polymerase. Hybridization was performed to *Eco*RI/*Hpa*II/*Bgl*II restriction fragments of the plasmid. The position of fragment A harbouring the rRNA promoter is indicated. Mc. th., *Methanococcus thermolithotrophicus*.

ment. With the exception of the G at the transcription start site, however, it is absent at the *hisA* gene.

The Methanococcus octanucleotide is a general element of archaeobacterial promoters

Comparison of the promoter element inferred from the binding studies with published DNA sequences revealed a high degree of homology at the same location between genes from both phylogenetic branches of archaeobacteria (Thomm and Wich 1988). Interestingly, the octanucleotide is well conserved in organisms with quite different GC contents of DNA. Stable RNA genes of *Thermoproteus tenax* (GC content of the DNA, 55%), for example, contain at the same location six to seven bases identical with the *Methanococcus* consensus. Even the promoters for stable RNA genes of extreme halophiles (GC content, 68%) show at -25 at least 5, and in most cases 6, nucleotides homologous to the AT-rich *Methanococcus* sequence (Thomm and Wich 1988). However, in position, 1 of the consensus T can be replaced by C, and the A in position 6 is usually substituted by a G (Fig. 2). This variation of the consensus depending on the GC content of the DNA was also observed in eubacteria (Bibb 1985); the Pribnow box of *Streptomyces*

promoters (GC content of the DNA, 70%) contains two G instead of a T and an A, as in the case of *E. coli* and *Bacillus* (GC content, 50%; Fig. 2).

The distance of the last nucleotide of the consensus to the ATGC motif varies between 16 and 23 bp (Fig. 2). A similar variation of the spacing of the conserved sequence to the transcription start site was observed in the case of the eukaryotic TATA box (Corden et al. 1980; Fig. 2).

Gene expression in archaeobacteria shows both eukaryotic and eubacterial features. The *Methanococcus* enzyme itself binds specifically to the *in vivo* transcription site as do the eubacterial holoenzymes (Fig. 2). The archaeobacterial enzyme covers a stretch of 50 bp corresponding to 17 nm of the B form of the DNA helix. A somewhat longer part of the DNA (70 or 60 bp) is protected by the RNA polymerases from *E. coli* (Siebenlist et al. 1980) and *Streptomyces* (Buttner and Brown 1985), respectively (Fig. 2).

Concerning the transcription signals, archaeobacterial promoters resemble much more eukaryotic polymerase II than do eubacterial promoters. Archaeobacteria share with eukaryotes a TATA box at -25 and a second conserved signal (although different in sequence) at the transcription start site (Fig. 2).

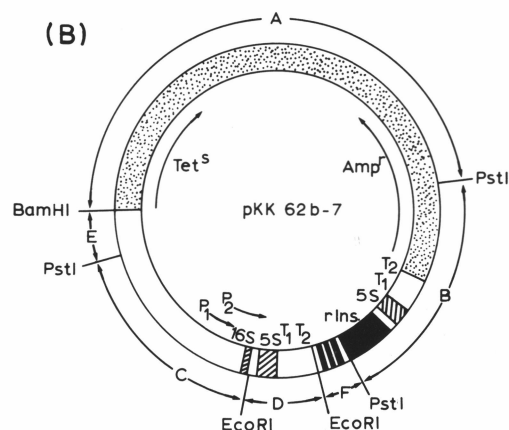
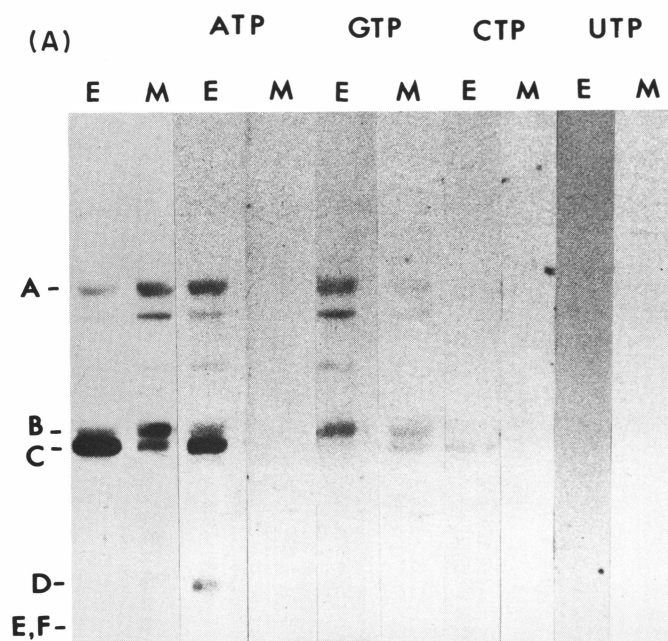


FIG. 4. (A) Southern hybridization of transcripts synthesized on a plasmid harbouring the promoters P1 and P2 of the *rrnB* operon of *E. coli* (Brosius 1984). E, *E. coli*; M, *Methanococcus thermolithotrophicus*. (B) *In vitro* RNA labelled with [α - 32 P]UTP (lanes 1 and 2) and γ - 32 P-ribonucleoside-5'-triphosphates (other lanes) was hybridized to a *Pst*I/*Bam*HI/*Eco*RI digest of plasmid pKK62b-7 (Brosius 1984). The promoters P1 and P2 are localized on DNA fragment C.

The absence of the -35 region and the different location of the AT-rich sequence clearly discriminate between archaeobacterial and eubacterial promoters.

On the basis of comparison of DNA sequences upstream of primary transcripts of *Sulfolobus* genes, a similar conclusion has been drawn by Reiter et al. (1988).

In vitro transcription experiments

Previously, we have shown by Southern hybridization experiments that the RNA polymerase of *M. thermolithotrophicus* initiates transcription selectively on a DNA segment harbouring the promoter of an rRNA operon (Thomm and Stetter 1985). To investigate specificity of transcription in more detail, the *in vitro* RNA synthesized on this template was end labelled with γ - 32 P-ribonucleoside-5'-triphosphates.

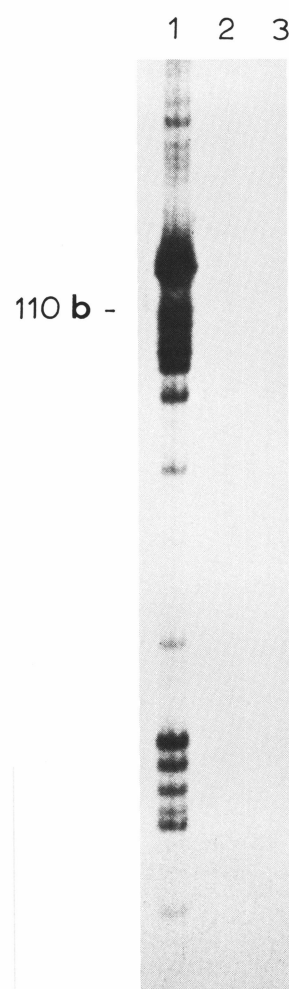


FIG. 5. Analysis of RNA products of different preparations of the *M. vannielii* RNA polymerase by electrophoresis in polyacrylamide urea gels. The relative activities of the different fractions on poly (dA-T) as a template are as follows: lane 1, 1.4; lane 2, 4.2; lane 3, 6.0. The amounts of protein and the levels of purification of the different preparations were identical.

Analysis of the *in vitro* RNA by Southern hybridization showed that only GTP was used as initiator nucleotide by the *Methanococcus* enzyme (Fig. 3). On the contrary RNAs synthesized on this template by the *E. coli* enzyme start with ATP, GTP, and CTP (Fig. 3).

To study, whether the specificity of the *Methanococcus* enzyme for GTP as start nucleotide is a general property of this enzyme, transcripts synthesized on a plasmid harbouring the promoters of a eubacterial rRNA operon were analysed. When the Southern hybridization was performed with transcripts labelled with [α - 32 P] UTP, only the *E. coli* RNA polymerase showed specificity for DNA fragment C, harbouring the promoters P1 and P2 of the *rrnB* operon (Fig. 4, left lane). Analysis of transcripts labelled with γ - 32 P-ribonucleoside-5'-triphosphates revealed that the rRNA transcripts of the *E. coli* enzyme start with GTP (at P1) and CTP (at P2) as has been described previously (Kingston and Chamberlin 1981; Fig. 4). The archaeobacterial enzyme, however, also on this template starts exclusively with GTP (Fig. 4), indicating that the specificity for GTP as initiator nucleotide is an intrinsic property of the *Methanococcus* enzyme. This might explain why the enzyme

can bind specifically to the *hisA* gene, although the ATGC motif at the transcription start site is absent there (Fig. 1). In the AT-rich intergenic region of *Methanococci*, a TATA box alone is certainly not sufficient to define a promoter. Considering the specificity of the enzyme for GTP as initiator nucleotide, the minimal requirement for a *Methanococcus* promoter might be a TATA box in a distance of about 25 bp to a C at the codogenic DNA strand.

Further analysis of *in vitro* transcripts synthesized by the *Methanococcus thermolithotrophicus* and *M. vannielii* RNA polymerase revealed that the purified enzymes are unable to produce correctly starting runoff transcripts (M. Thomm and G. Frey, unpublished). Although the *M. vannielii* RNA polymerase purified by a standard procedure (Thomm and Wich 1988; Brown et al. 1988) binds specifically to the promoter and shows high activity with poly (dA-T) as a template (Fig. 5), these enzyme fractions usually do not initiate transcripts on supercoiled templates (Fig. 5, lanes 2 and 3). Some preparations (for unknown reasons) of the enzyme, however, are able to synthesize defined transcripts on a plasmid harbouring the tRNA^{val} of *M. vannielii* (Fig. 5, lane 1). One of the RNAs shows a length of 110 bases. This transcript has the length expected for a primary tRNA transcript starting and terminating at the *in vivo* sites (Wich et al. 1986). This might be considered as an indication that a transcription factor is present in this RNA polymerase preparation, which is necessary for initiation of transcription. A modified purification procedure revealed indeed in a reproducible manner an RNA polymerase fraction that initiates and terminates transcription correctly at this template (M. Thomm, G. Frey, and W. Hausner, in preparation).

The fact that the archaeobacterial RNA polymerase is able to bind specifically to the promoter and seems to require a soluble factor for initiation of transcription suggests a novel function of the transcription factor in archaeobacteria.

Acknowledgements

This work was supported by a grant of the Deutsche Forschungsgemeinschaft and the North Atlantic Treaty Organization Collaborative Research Grant No. 0148/85.

- BIBB, M. J. 1985. Gene expression in *Streptomyces*—nucleotide sequences involved in the initiation of transcription and translation. In *Biological, biochemical and biomedical aspects of Actinomycetes*. Edited by G. Szabo, S. Biro, and M. Goodfellow. Akademia Kiado, Budapest. pp. 25–35.
- BROSIOUS, J. 1984. Toxicity of an overproduced foreign gene product in *Escherichia coli* and its use in plasmid vectors for the selection of transcription terminators. *Gene*, **27**: 161–172.
- BROWN, J. W., THOMM, M., BECKLER, G. S., FREY, G., STETTER, K. O., and REEVE, J. 1988. An archaeobacterial RNA polymerase binding site and transcription initiation of the *hisA* gene of *Methanococcus vannielii*. *Nucleic Acids Res.* **16**: 135–150.
- BURGESS, R. R., TRAVERS, A. A., DUNN, J. J., and BAUTZ, E. K. F. 1969. Factor stimulating transcription of RNA polymerase. *Nature (London)*, **221**: 43–46.
- BUTTNER, M. J., and BROWN, N. L. 1985. RNA polymerase – DNA interactions in *Streptomyces*. *J. Mol. Biol.* **185**: 177–188.
- CORDEN, J., WASYLYK, B., BUCHWALDER, A., SASSONE-CORSI, P., KEDINGER, C., and CHAMBON, P. 1980. Promoter sequences of eukaryotic protein coding genes. *Science (Washington, D.C.)*, **209**: 1406–1414.
- DYNAN, W. S., and TJIAN, R. 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature (London)*, **316**: 774–778.
- GALAS, D. J., and SCHMITZ, A. 1978. DNase footprinting a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5**: 3157–3170.
- HINKLE, D. C., and CHAMBERLIN, M. J. 1972. Studies of the binding of *Escherichia coli* RNA polymerase to DNA. I. The role of sigma subunit site selection. *J. Mol. Biol.* **70**: 157–185.
- HOOPES, B. C., and MCCLURE, W. R. 1987. Strategies in regulation of transcription initiation. In *Escherichia coli and Salmonella typhimurium cellular and molecular biology*. Edited by J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger. American Society for Microbiology, Washington, DC. pp. 1231–1240.
- JACOB, F., ULLMAN, A., and MONOD, J. 1964. Le promoteur, élément génétique nécessaire à l'expression d'un operon. *C. R. Seances Acad. Sci. Ser. 3*, **258**: 3125–3128.
- KINGSTON, R. E., and CHAMBERLIN, M. 1981. Pausing and attenuation of *in vitro* transcription in the *rrmB* operon of *E. coli*. *Cell*, **27**: 523–531.
- PRIBNOW, D. 1975. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc. Natl. Acad. Sci. U.S.A.* **72**: 784–788.
- REITER, W. D., PALM, P., and ZILLIG, W. 1988. Analysis of transcription in the archaeobacterium *Sulfolobus* indicates that archaeobacterial promoters are homologous to eukaryotic pol II promoters. *Nucleic Acids Res.* **16**: 1–19.
- SCHALLER, H., GRAY, L., and HERRMANN, K. 1975. Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd. *Proc. Natl. Acad. Sci. U.S.A.* **72**: 737–741.
- SHALLOWAY, D., KLEINBERGER, T., and LIVINGSTON, D. M. 1980. Mapping of SV40 DNA replication origin binding sites for the SV40 T antigen by protection against exonuclease III digestion. *Cell*, **20**: 411–422.
- SIEBENLIST, U., SIMPSON, R. B., and GILBERT, W. 1980. *E. coli* RNA polymerase interacts homologously with two different promoters. *Cell*, **20**: 269–281.
- THOMM, M., and STETTER, K. O. 1985. Transcription in methanogens. Evidence for specific *in vitro* transcription of the purified DNA-dependent RNA polymerase of *Methanococcus thermolithotrophicus*. *Eur. J. Biochem.* **149**: 345–351.
- THOMM, M., and WICH, G. 1988. An archaeobacterial promoter element for stable RNA genes with homology to the TATA-box of higher eukaryotes. *Nucleic Acids Res.* **16**: 151–163.
- THOMM, M., SHERF, B., and REEVE, J. N. 1988. RNA polymerase binding and transcription initiation sites upstream of the methyl reductase operon of *Methanococcus vannielii*. *J. Bacteriol.* **170**: 1958–1961.
- WICH, G., HUMMEL, H., JARSCH, M., BÄR, U., and BÖCK, A. 1986. Transcription signals for stable RNA genes in *Methanococcus*. *Nucleic Acids Res.* **14**: 2459–2479.
- WOESE, C. R., and FOX, G. E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* **74**: 5088–5090.
- ZILLIG, W., STETTER, K. O., SCHNABEL, R., MADON, J., and GIERL, A. 1982. Transcription in archaeobacteria. *Zentralbl. Bakteriol. Mikrobiol. Hyg. Abt. 1 Orig. C3*: 218–227.